Balance and spontaneous wave generation in geophysical flows

J. Vanneste

School of Mathematics and Maxwell Institute for Mathematical Sciences,

University of Edinburgh, Edinburgh EH9 3JZ, UK

J.Vanneste@ed.ac.uk

Key Words inertia-gravity wave, Rossby number, slow manifold, exponential asymptotics, quasi-geostrophic model

Abstract The large-scale dynamics of the mid-latitude atmosphere and ocean is characterised by a time-scale separation between slow balanced motion and fast inertia-gravity waves. As a result of this separation, the two types of motion interact only weakly, and the dynamics can be approximated using balanced models which filter out the fast waves completely. The separation is not complete, however: the evolution of well-balanced flows inevitably leads to the excitation of inertia-gravity waves through the process of spontaneous generation. Spontaneous generation has fundamental and practical implications: it limits the validity of balanced models, and provides a source of inertia-gravity-wave activity. These two aspects are discussed in this review, which focusses on the small-Rossby-number regime $\epsilon \ll 1$ corresponding to strong rotation. Theoretical arguments indicate that spontaneous generation is then exponentially small in ϵ for smooth flows. They are complemented by numerical simulations which identify specific generation mechanisms.

CONTENTS

1 Introduction

The circulation of the atmosphere and ocean at mid-latitudes is dominated by large-scale motion evolving over long time scales, of the order of a few days in the atmosphere and a few weeks in the ocean. This slow motion is termed balanced motion because the pressure gradients are nearly balanced by the Coriolis force in the horizontal (geostrophic balance) and by buoyancy forces in the vertical (hydrostatic balance). Although dominant in much of the atmosphere and ocean, this type of motion is not the only possible one: small-scale, fast inertia-gravity

waves (IGWs), with frequencies greater than or equal to the Coriolis frequency f, provide an additional mode of motion which is observed routinely. These waves, with time scales ranging from minutes to hours, play an important role, for instance by transporting momentum vertically and by enhancing mixing. Their main sources are topography and convection in the atmosphere, wind and tidal forcing in the ocean.

The time-scale separation between balanced motion and IGWs is estimated by the Rossby number

$$\epsilon = \frac{U}{fL}, \tag{1}$$

where U and L are relevant horizontal velocity and length scales. For large-scale flows, it takes typical values of the order of 0.1 or smaller (e.g. Vallis 2006). These flows are therefore in the rapidly rotating, small-Rossby-number regime $\epsilon \ll 1$ characterised by a large time-scale separation. Because of this, the coupling between balanced motion and IGWs is weak, and balanced models (such as the quasi-geostrophic (QG) model), which filter out IGWs entirely, capture much of the dynamics.

The coupling between balanced and IGW motion cannot be completely ignored, however, and a long-standing issue in geophysical fluid dynamics has been its quantification. There are several reasons for the importance given to this issue. From a theoretical viewpoint, the coupling between balanced motion and IGWs is an obstacle to the existence of an exactly invariant slow manifold, the idealisation put forward by Leith (1980) and Lorenz (1980) of a subset of phase space on which the dynamics is completely devoid of IGW activity. Correspondingly, it places a fundamental limitation on the accuracy of balanced models. At a more practical level, the coupling has consequences for the initialisation schemes (e.g.

Daley 1991) used in operational meteorology to filter out IGWs, most of which spurious, from initial data derived from observations. Importantly, it is also at the root of spontaneous generation, or spontaneous adjustment, the mechanism whereby slow, well-balanced motion emits IGWs in the course of its evolution. This mechanism should be contrasted with standard adjustment (Rossby 1937) in which IGW generation results from an initial imbalance imposed externally. It is an additional source of IGWs which, unlike the sources mentioned above, can potentially excite waves across the whole spectrum of frequencies and wavelengths.

While the issues surrounding the coupling between balanced motion and IGWs have attracted continuous attention from the 1980s onward (see sections 2.1–2.2 below), recent developments have considerably improved our understanding. They are the main subject of this review. One of these is the connection with dynamical-system work on non-dissipative two-time-scale systems which indicates that the coupling is exponentially small for $\epsilon \ll 1$ (sections 2.3–2.4). Another is a series of high-resolution numerical simulations of spontaneous generation in different flow configurations (section 3). These identify specific generation mechanisms and demonstrate convincingly their relevance to atmospheric and oceanic flows. A third strand of development consists in the asymptotic analysis of some highly idealised flow models which yields explicit estimates for the exponentially small amplitudes (or growth rates) of IGWs generated spontaneously (section 4).

This review focusses on the small-Rossby-number regime $\epsilon \ll 1$ and on mechanisms of wave generation that apply to well-balanced (in a sense made precise) initial conditions. For completeness, we nonetheless briefly discuss averaging techniques relevant to flows with initial imbalance (section 2.5), as well as the

strong-stratification, small-Froude-number regime in which Lighthill radiation provides alternative mechanisms of wave generation (section 5).

2 Balance and its breakdown

2.1 Primitive equations

To fix ideas, let us consider the primitive equations (PEs)[1] on the f-plane in the Boussinesq approximation (e.g. Vallis 2006), written as

$$\frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}t} + f\hat{\boldsymbol{z}}\times\boldsymbol{u} = -\nabla\phi + b\hat{\boldsymbol{z}}, \tag{2a}$$

$$\frac{\mathrm{d}b}{\mathrm{d}t} + N^2 w = 0, \tag{2b}$$

$$\nabla\cdot\boldsymbol{u} = 0, \tag{2c}$$

where $\boldsymbol{u}=(u,v,w)$ is the velocity, $\mathrm{d}/\mathrm{d}t=\partial_t+\boldsymbol{u}\cdot\nabla$ is the material derivative, f is the Coriolis frequency, ϕ is a scaled pressure, and $\hat{\boldsymbol{z}}$ is the vertical unit vector. The buoyancy has been written as N^2z+b, with a Brunt–Väisälä frequency N that is assumed constant for simplicity. The form (2) obscures the important point that the dynamics is in fact governed by the evolution of only 3 independent fields. Different choices can made for these fields; it is convenient to take the (perturbation) potential vorticity (PV)

$$q = (f\hat{\boldsymbol{z}}+\nabla\times\boldsymbol{u})\cdot\big(N^2\hat{\boldsymbol{z}}+\nabla b\big) - fN^2, \tag{3}$$

the horizontal divergence $\delta=u_x+v_y$, and the 'ageostrophic vorticity' $\gamma=f(v_x-u_y)-\nabla_{\mathrm{h}}^2\phi$, where ∇_{h}^2 is the horizontal Laplacian (e.g. Mohebalhojeh & Dritschel

[1] We use the term PEs in a broader sense than is usual to refer to equations filtering acoustic waves but not inertia-gravity waves, without implying that the hydrostatic approximation is made.

2001). In terms of these, (2) become

$$\partial_t q = -\boldsymbol{u}\cdot\nabla q, \tag{4a}$$

$$\partial_t \delta - \gamma = N_\delta, \tag{4b}$$

$$\partial_t \gamma + \mathcal{K}\delta = N_\gamma, \tag{4c}$$

where N_δ and N_γ group nonlinear terms, and $\mathcal{K}$ is the linear operator

$$\mathcal{K} = f^2 + (N^2 - f^2)\nabla_{\mathrm{h}}^2\nabla^{-2}.$$

Eqs. (4) form a closed system since $(\boldsymbol{u}, \phi, b)$ appearing on the right-hand sides are recovered from (q, δ, γ) by inverting nonlinear elliptic operators. These inversions involve boundary conditions. For domains bounded in the vertical by rigid horizontal lids at $z = 0$, H, these require to solve

$$\partial_t b = -\boldsymbol{u}\cdot\nabla b, \quad \text{at} \quad z = 0,\, H, \tag{5}$$

which follows from (2b) and the condition $w = 0$ at $z = 0$, H. Thus the PEs have the two dimensional fields $b(x, y, 0, t)$ and $b(x, y, H, t)$ as additional degrees of freedom. Further complications arise in the horizontal, especially for multiply connected domains (cf. Yuan & Hamilton 1994, Muraki et al. 1999).

Linearising (4) and assuming normal-mode solutions $(q, \delta, \gamma) \propto \mathrm{e}^{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x} - \omega t)}$, with $\boldsymbol{k} = (k, l, m)$ a wavevector, leads to the three branches of the dispersion relation

$$\omega_0 = 0 \quad \text{and} \quad \omega_\pm = \pm\left(f^2 + \frac{(N^2 - f^2)\kappa^2}{\kappa^2 + m^2}\right)^{1/2}, \tag{6}$$

where $\kappa = (k^2 + l^2)^{1/2}$ is the horizontal wavenumber. The first branch corresponds to the vortical, or balanced mode; the other two branches correspond to IGWs propagating in the directions of $\pm\boldsymbol{k}$ (e.g. Gill 1982). For $N > f$ (typically $N \gg f$), the IGW frequencies are bounded from below by f: $|\omega_\pm| \geq f$.

The balanced mode describes the dynamics of PV which is trivial at a linear level but in fact governed by nonlinear advection. A relevant frequency for this part of the dynamics is therefore $\omega_0 \approx U/L$, where U and L are characteristic horizontal velocity and length scales, rather than 0. This leads to the ratio between the frequencies of the vortical and balanced modes

$$\begin{aligned} \frac{\omega_0}{\omega_\pm} &\approx \frac{U}{L\left(f^2 + (N^2 - f^2)\kappa^2/(\kappa^2 + m^2)\right)^{1/2}} \\ &\approx \frac{\epsilon}{\left(1 + (s^2 - 1)\kappa^2/(\kappa^2 + m^2)\right)^{1/2}} \leq \epsilon, \end{aligned} \tag{7}$$

where the Rossby number ϵ is defined in (1) and $s = N/f$ is assumed to satisfy $s > 1$. In the atmosphere, typical values are $U \approx 10\,\mathrm{m\,s^{-1}}$, $L \approx 1000\,\mathrm{km}$ and $f \approx 10^{-4}\,\mathrm{s^{-1}}$, yielding $\epsilon \approx 0.1$; in the ocean, typical values of ϵ are smaller still. Thus, at large scales, the atmosphere and ocean are in the rapidly rotating, small-Rossby-number regime, $\epsilon \ll 1$, in which $|\omega_\pm| \gg |\omega_0|$.

The regime $\epsilon \ll 1$ is not the only one in which the frequencies are widely separated. Another appears when $\epsilon = O(1)$ for strong stratification, corresponding to a Froude number

$$F = \frac{U}{NH} \ll 1, \tag{8}$$

where H is a vertical scale. Saujani & Shepherd (2006) propose to use the ratio (7) as small parameter to capture the regimes $\epsilon \ll 1$ and $F \ll 1$ in a unified way. There is however a crucial difference between the two regimes when it comes to spontaneous wave generation. This appears because the spatial scales of the balanced motion and of the waves are not necessarily the same, as assumed so far. The condition $\epsilon \ll 1$ guarantees a frequency separation between between balanced motion and waves regardless of their relative scales. In contrast, the condition $F \ll 1$ guarantees frequency separation only for waves with $\kappa L/(mH) = O(1)$;

waves with shallow enough structure that $\kappa L/(mH) = O(F)$ have low frequencies that match those of the balanced motion; this is central to the mechanism of Lighthill radiation discussed in section5. The rest of this review focuses on the small-Rossby-number regime $\epsilon \ll 1$.

2.2 Slow manifold, balance relations and balanced models

Using L/U to non-dimensionalise t, the PEs (4) can be written in the abstract form

$$\frac{\partial \mathsf{s}}{\partial t} = \mathsf{N_s}(\mathsf{s}, \mathsf{f}), \tag{9a}$$

$$\frac{\partial \mathsf{f}}{\partial t} + \frac{1}{\epsilon}\mathcal{L}\mathsf{f} = \mathsf{N_f}(\mathsf{s}, \mathsf{f}), \tag{9b}$$

with $\mathsf{s} = q$ and $\mathsf{f} = (\delta, \gamma/f)$,

$$\mathcal{L} = \begin{pmatrix} 0 & -1 \\ 1 + (s^2 - 1)\nabla_{\mathrm{h}}^2\nabla^{-2} & 0 \end{pmatrix}$$

and the nonlinear terms are again grouped in the right-hand sides. The shallow-water model, often used instead of the more complicated PEs, also fits this form.

It has proved illuminating to study the problem of balance and wave generation by considering the general class of two-time-scale systems described by (9) (Warn et al. 1995): one advantage is that essential properties can be examined unencumbered by the details of specific models; another is that insight is gained from very simple finite-dimensional models fitting the form (9). The defining features of the abstract model are two properties of the spectrum of the linear operator $\mathcal{L}$: it is (i) purely imaginary and (ii) bounded from below by 1. The first property, which implies that the linear dynamics of f is oscillatory, reflects the non-dissipative nature of large-scale geophysical fluids; the second ensures the complete frequency separation between slow and fast variables.

LK model: Lorenz obtained this model by truncation of the shallow-water equations, retaining only the amplitudes $\mathsf{s} = (u, v, w)$ of 3 vortical modes as slow variables, and the amplitudes of a gravity-wave pair $\mathsf{f} = (x, y)$ as fast variables. These amplitudes obey the 5 ODEs

$$\dot{u} = -vw + \epsilon bvy, \quad \dot{v} = uw - \epsilon buy, \quad \dot{w} = -uv,$$

$$\epsilon\dot{x} = -y, \quad \epsilon\dot{y} = x + buv,$$

where $\epsilon = U/(fL\sqrt{1+b^2})$ is the frequency-separation parameter analogous to (7) and $b = \sqrt{gH}/(fL)$. The substitution $(x, y) = C(\cos(\theta/2), \sin(\theta/2))$, where C is a constant, leads to ODEs very similar to those describing the elastic pendulum of Fig. 1.

Lorenz pioneered the use of finite-dimensional models when he proposed his 5-component model (Lorenz 1980), also referred to as Lorenz–Krishnamurthy (LK) model after a subsequent work (Lorenz & Krishnamurthy 1987). See sidebar for details. This model has been studied extensively (Jacobs 1991, Lorenz 1992, Boyd 1994, Fowler & Kember 1996, Camassa 1995, Bokhove & Shepherd 1996, Camassa & Tin 1996). Camassa (1995) and Bokhove & Shepherd (1996) pointed out that it reduces to a two-degree of freedom Hamiltonian system very similar to a simple mechanical system, namely the elastic pendulum, or swinging spring. This consists of a mass attached to a fixed point by a stiff spring which moves in the vertical plane under the combined action of gravity and spring restoring force (Lynch 2002) (see Fig. 1). The small parameter in this case is the ratio of the pendulum frequency to the spring frequency. Like the LK model, the elastic pendulum provides a simple, intuitive model with which to explore balance and spontaneous wave generation: the pendulum oscillations are analogous to the

slow, balanced motion while the fast vibrations are analogous to IGWs.

In two-time-scale systems, the interactions between the slow and fast degrees of freedom are weak. If, as is the case in the atmosphere and ocean, the forcing is predominantly slow, the fast dynamics is only weakly excited. The central idea of balance is to seek the conditions that minimize this excitation. In the first approximation, this is achieved by setting $\mathsf{f} = 0$, that is, for the PEs (4), $\delta = \gamma = 0$, corresponding to geostrophic and hydrostatic balance,

$$u = -f\phi_y, \quad v = f\phi_x, \quad b = \phi_z.$$

This is a first example of a *balance relation* relating the fast variables to the slow ones. Numerous balance relations improving on this have been proposed. They are best thought of as defining *slow manifolds* (Leith 1980, Lorenz 1980). These are manifolds (with dimension $\dim \mathsf{s}$) in the state space of the system of the form

$$\mathsf{f} = \mathsf{F}(\mathsf{s}, \epsilon), \tag{10}$$

with $\mathsf{F}(\mathsf{s}, 0) = 0$, which are nearly invariant and on which the dynamics is slow (Warn et al. 1995, MacKay 2004). Nearly invariant means that trajectories of the full system starting on the manifold stay close to it; this can be measured by the angle between the vector field $(\partial_t \mathsf{s}, \partial_t \mathsf{f})$ and the slow manifold. Fig. 2 gives a pictorial representation of a slow manifold, with an indication of an exact trajectory and its balanced projection. Much of the geophysical literature refers to *the* slow manifold, requiring implicitly that the manifold be exactly invariant. However, since it has become clear that no such invariant object exists for non-dissipative systems (MacKay 2004), it is preferable to conform to the dynamical-systems usage and consider a hierarchy of slow manifolds, or equivalently balance relations, of increasing accuracy.

The derivation of increasingly accurate balance relations can be made systematic by regarding these as approximate solutions of the 'superbalance equation'

$$\epsilon \mathsf{N_s}(\mathsf{s},\mathsf{F}) \cdot \partial_{\mathsf{s}} \mathsf{F} + \mathcal{L}\mathsf{F} = \epsilon \mathsf{N_f}(\mathsf{s},\mathsf{F}) \tag{11}$$

obtained by introducing (10) into (9). Introducing a power-series expansion of the unknown function $\mathsf{F}(\mathsf{s},\epsilon)$,

$$\mathsf{F}(\mathsf{s},\epsilon) = \sum_{n=1}^{N} \epsilon^n \mathsf{F}^{(n)}(\mathsf{s}), \tag{12}$$

leads to one class of balanced models (Warn et al. 1995); using iterations such as

$$\epsilon \mathsf{N_s}(\mathsf{s},\mathsf{F}^{(n)}) \cdot \partial_{\mathsf{s}} \mathsf{F}^{(n)} + \mathcal{L}\mathsf{F}^{(n+1)} = \epsilon \mathsf{N_f}(\mathsf{s},\mathsf{F}^{(n)}), \tag{13}$$

where $\mathsf{F}^{(n)}$ now denotes the iterate, leads to another (Allen 1993, Warn et al. 1995). Still other balance relations can be derived using the method of bounded derivatives (Kreiss 1991, and references therein) which sets to 0 some derivative $\partial_t^N \mathsf{f}$ of the fast variables (see McIntyre & Norton (2000) and Mohebalhojeh & Dritschel (2001) for numerical implementation of these relations as well as for references to their history). The precise form of the balance relations further depends on the choice of the slow and fast variables or, more generally, on the way the slow manifolds are parameterised.

For a given balance relation $\mathsf{F}(\mathsf{s},\epsilon)$, that is, for a given slow manifold, the full dynamics can be approximated by projection on the slow manifold. This leads to a reduced *balanced model*, which filters out fast oscillations. The simplest projection fixes s to yield the balanced model

$$\frac{\partial \mathsf{s}}{\partial t} = \mathsf{N_s}(\mathsf{s},\mathsf{F}(\mathsf{s},\epsilon)). \tag{14}$$

For the PEs with geostrophic balance, this corresponds to the standard QG model (e.g. Vallis 2006). More accurate balance relations lead to more accurate balanced

models. There is a great deal of freedom in the choice of the projection, which can be exploited to obtain balanced models satisfying desirable properties: the preservation of the Hamiltonian structure of the PEs has attracted a great deal of attention (Salmon 1983, 1985, McIntyre & Roulstone 2002); others properties are the conservation of PV and mass (Mohebalhojeh & McIntyre 2007), the well-posedness of the balanced model, and its validity in multiple regimes.

We do not review here the literature devoted to the derivation and applications of the huge variety of balanced models issued from different balance relations and projections. We simply note that, from an asymptotic viewpoint, all balance relations and balanced models of a given accuracy are equivalent, differing only by terms that are smaller than the error made in approximating the superbalance equations.

2.3 Breakdown of balance and spontaneous generation

The accuracy of balance relations can be improved order-by-order in ϵ without any apparent obstacles. A natural question, therefore, is whether an exact solution of the superbalance relation, corresponding to an exactly invariant slow manifold, can be found. This question, for some time controversial, has generated a great deal of activity (e.g. Lorenz 1980, Vautard & Legras 1986, Warn & Menard 1986, Lorenz 1986, 1992, Warn 1997). For the elastic pendulum, it has an intuitive interpretation: given the initial angle and angular velocity of the pendulum, is it possible to find an initial spring stretch and velocity such that fast oscillations of the spring are never excited subsequently.

The answer to this question can be found in the dynamical-systems literature and hinges on the non-dissipative nature of the relevant models: whereas for dis-

sipative systems, exactly invariant, attracting slow manifolds can be obtained, this is not the case for non-dissipative systems. For (generic) non-dissipative systems, there is no exactly invariant slow manifold. The series-expansion or iteration procedures used to compute slow manifolds can be carried out to arbitrary accuracy $O(\epsilon^N)$, but they do not converge as $N \to \infty$. These procedures are only asymptotic: for fixed N, the error decreases as $\epsilon \to 0$; but for fixed ϵ, the error decreases as N increases up to some optimal value N_ϵ, then increases (e.g. Bender & Orszag 1999). The divergence is not the product of particular asymptotic procedures; rather it is a consequence of the fundamental property of non-existence of an invariant manifold.

Using the asymptotic nature of the power series or iterations used to construct slow manifolds, it is possible to put an upper bound on the amplitude of the fast motion. A typical behaviour for the coefficients $\mathsf{F}^{(n)}$ in (12) is factorial-like growth: $\mathsf{F}^{(n)} \propto n!$ as $n \to \infty$. A slow manifold with near-optimal accuracy is obtained by optimal truncation, that is by choosing the number of terms N in the series as the number N_ϵ which minimises the last term $\epsilon^N \mathsf{F}^{(N)} \propto \epsilon^N N!$. It follows that $N_\epsilon \propto \epsilon^{-1}$, leading to an accuracy that is exponential in ϵ, of the form $\exp(-\alpha/\epsilon)$ for some $\alpha > 0$ (multiplied by a prefactor $\gamma\epsilon^\beta$ for some β and γ).

The non-existence of an exactly invariant slow manifold means that, however well initialised, the slow evolution of balanced motion is always accompanied by fast motion; in other words spontaneous generation — of fast oscillations for the elastic pendulum, of IGWs for geophysical flows — is inevitable, as represented schematically in Fig. 3. This led Warn (1997) to introduce the notion of *fuzzy slow manifold* to describe, in a loose way, the narrow region surrounding the most accurate slow manifolds. The optimal-truncation argument sketched above

indicates that the spontaneous generation of IGW is exponentially small in ϵ; in other words, the width of the fuzzy manifold is exponentially small. This point was made by Lorenz & Krishnamurthy (1987), on the basis of numerical solutions of the LK model, and by Warn (1997). It can be understood heuristically by interpreting wave generation as caused by the resonant excitation of fast motion by the high-frequency component of the slow motion; this high-frequency component is exponentially small if the slow motion is analytic in t and hence has a power spectrum that decays exponentially at large frequencies.

The heuristic arguments above can be made rigorous using a method of Nekhoroshev (1977) and Neishtadt (1982) which yields exponentially small bounds on the accuracy of slow manifolds obtained by optimally truncated iterations (see also Gelfreich & Lerman 2002, MacKay 2004). This method has been applied for simple geophysical models by Cotter (2004), Cotter & Reich (2006) and Wirosoetisno (2004). More recently, Temam & Wirosoetisno (2007, 2010, 2011) considered the hydrostatic PEs including viscous terms and obtained a bound of the form $\exp(-\alpha/\epsilon^{1/3})$ using a method adapted to dissipative PDEs (Matthies 2001). The bound in this case relies on dissipation and increases rapidly with the Reynolds number; this limits its direct relevance to geophysical systems, where the Rossby number ϵ is much larger than the inverse Reynolds number.

2.4 Beyond all orders

The fact that spontaneous generation is exponentially small in ϵ does not mean that it is negligible or impossible to compute. To illustrate the first point, Fig. 4 shows the evolution of one of the fast variables in the LK model for an initial condition analogous to that of a pendulum starting near its upright position. Because

the corresponding balanced motion is asymptotic to a constant as $t \to \pm\infty$, the identification of fast oscillations is straightforward: clearly, the evolution leads to the excitation of oscillations around $t = 0$. The amplitude of these oscillations decreases sharply with ϵ, as expected from the exponential dependence, but it is significant compared with the changes associated with the balanced evolution for ϵ moderately small.

The second point is that it is possible to derive asymptotic estimates for the fast oscillations generated spontaneously in a model such as the LK model using the techniques of exponential asymptotics. Several approaches are available to capture the exponentially small terms (e.g. Segur et al. 1991, Olde Daalhuis et al. 1995, Balser 2000, Hakim 1991). These approaches recognise that the total solution is the sum of a dominant, balanced contribution, given by a first asymptotic series, and a subdominant contribution representing fast oscillations and given by another series. Spontaneous generation corresponds to the switching on of the second series by the first, and is an instance of the Stokes phenomenon (e.g. Ablowitz & Fokas 1997).

To understand the origin of the switching on, let us consider a single solution starting near a slow manifold. The fast variables can be written as

$$\mathsf{f}(t) = \sum_{n=1}^{N} \epsilon^n \mathsf{f}_{\rm bal}^{(n)}(t) + \mathsf{f}_{\rm rem}^{(N)}(t), \tag{15}$$

where the terms in the series, obtained by straightfoward expansion, represent the balanced part of the solution. The remainder $\mathsf{f}_{\rm rem}^{(N)}(t)$ satisfies an equation of the form

$$\epsilon \partial_t \mathsf{f}_{\rm rem}^{(N)} + \mathcal{L}\mathsf{f}_{\rm rem}^{(N)} = -\epsilon^{N+1}\left(\partial_t \mathsf{f}_{\rm bal}^{(N)} + \cdots\right) \tag{16}$$

where $\cdots$ groups nonlinear terms whose details are unimportant for our argument. For $N = O(1)$, the first term is negligible, and the remainder can be approximated

as $\epsilon^{N+1}\mathcal{L}^{-1}\left(\partial_t \mathsf{f}_{\rm bal}^{(N)} + \cdots\right) = \epsilon^{N+1}\mathsf{f}_{\rm bal}^{(N+1)}$. This amounts to the addition of an extra term to the asymptotic series in (15). The process cannot be continued for arbitrary large N, however. Because of repeated time differentiation, the right-hand side grows: roughly, for smooth solutions, $\mathsf{f}^{(N)} \propto \partial_t^N \mathsf{f}^{(0)} \propto N!/(t-t_*)^N$, where t_* is the complex-time singularity of $\mathsf{f}^{(0)}$ nearest the real line (Berry 2005). Near optimal truncation, say $N \sim N_\epsilon \propto \epsilon^{-1}$, the right-hand side is exponentially small but a rapidly varying function for t near $\tau = \mathrm{Re}\, t_*$. As a result, the time derivative on the left-hand side of (16) cannot be neglected, and the remainder $\mathsf{f}_{\rm rem}^{(N)}$ is dominated by rapid oscillations.

The techniques of exponential asymptotics provide approximations for $\mathsf{f}_{\rm rem}^{(N)}$ and hence for the oscillations generated spontaneously; they also reveal a number of qualitative features. First, the generation — the switching on of the subdominant series — occurs at the specific time $t = \tau = \mathrm{Re}\, t_*$ corresponding to the intersection of the real t-axis with the line, termed Stokes line, joining the singularities t_* and $\bar{t}_*$ of the balanced motion. This explains the sudden generation at $t = 0$ observed for the solutions in Fig. 4 for which $\tau = 0$. Moreover, the switching on occurs on an $O(\epsilon^{1/2})$ time scale and is generically described by an error function (Berry 1989).

Second, it emerges that t_* controls the constant α in the asymptotics $\exp(-\alpha/\epsilon)$ of spontaneous generation: in the simplest cases, $\alpha = |\omega_\pm \mathrm{Im}\, t_*|$, where $\omega_\pm$ is the scaled IGW frequency. The connection is not surprising: heuristically, the oscillation amplitude depends on the content of the balanced signal at the IGW frequency $\omega_\pm/\epsilon$; this is precisely given by $\exp(-|\omega_\pm \mathrm{Im}\, t_*|/\epsilon)$.

Finally, the exponential-asymptotics techniques rely on an approximation to $\mathsf{f}_{\rm bal}^{(n)}$ for $n \gg 1$ (the so-called late terms of (15)). This implies that the expo-

nentially small terms cannot be estimated from the leading-order approximation (QG in the geophysical context) to the balanced motion only but require a much more detailed understanding of this motion. This contrasts with the situation in the small-Froude-number regime (see section 5) and suggests that an analytic description of spontaneous generation will only be practical for highly simplified models. Such a description has been obtained for the LK model (Vanneste 2004) and for simple solutions of the fluid equations (see section 4 below). The results for the LK model are illustrated by Fig. 5 which shows an estimate of the unbalanced part of the fast variable x derived from a highly accurate balance relation. The fast oscillations are clearly switched on around the time $t = \tau = 0$ and are in agreement with asymptotic predictions.

It may be noted that knowledge of the asymptotic behaviour of the late terms, $\mathsf{f}_{\rm bal}^{(n)}$ for $n \gg 1$, makes it possible to apply a Borel-resummation technique (e.g. Hinch 1991) in order to give a finite meaning to the divergent series representing the balanced motion. In principle at least, this allows to define a unique slow manifold which is essentially invariant except across Stokes lines where spontaneous generation takes place (Vanneste 2008a). This is illustrated schematically in Fig. 3.

The discussion in this section emphasises the temporal aspects of spontaneous generation as relevant to ODE models; for PDEs, however, one also needs to consider spatial aspects. In particular, the asymptotic series defining slow manifolds for PDEs involve repeated spatial differentiations, leading to a divergence associated with singularities of the balanced motion in (complex) space. Much less is known about these aspects, although the approach of Temam & Wirosoetisno (2007, 2010, 2011), which relies on a spectral expansion of the spatial depen-

dence truncated in an ϵ-dependent manner, provides a way of establishing the exponential accuracy of balance for PDEs.

2.5 Averaging

Spontaneous generation of IGWs arises in flows that are assumed to be well balanced. This assumption is motivated to some extent by the low level of IGW activity in large parts of the atmosphere, but also by the links with initialisation procedures and by the pragmatic view taken that IGWs are so poorly constrained by observations that they are better neglected. However, the time-scale separation between balanced motion and IGWs leads to interactions between the two types of motion that are weak even for IGWs with $O(1)$ amplitudes. This is captured by averaging the PEs over the fast IGWs time scale to obtain slow equations which govern both the balanced motion and the amplitude (but not the phase) of IGWs (Embid & Majda 1996, 1998, Majda & Embid 1998, Babin et al. 2000, 2001, 2002, Wirosoetisno et al. 2002). These equations, which retain only resonant interactions, have an interesting structure: first, to leading-order, the balanced dynamics is governed by the QG equations just as it is near a slow manifold. This is a consequence of the conservation of PV and the resulting vanishing of the quadratic forcing of balanced modes by IGWs (Warn 1986). Second, IGW interactions are restricted to waves with the same frequency catalysed by the balanced motion. Third, the behaviour of IGWs differs qualitatively in bounded and unbounded domains. In the latter case, the rapid dispersion of IGWs to infinity further limits their possible effect on the balanced motion (Reznik et al. 2001, Zeitlin et al. 2003).

3 Numerical simulations

In recent years, a number of high-resolution numerical simulations of the PEs have demonstrated the relevance of spontaneous IGW generation to realistic atmospheric and oceanic flows. Such simulations are delicate because of the small amplitudes and often small scales of the waves generated. Not surprisingly, most attention has been paid to flows in which ϵ is order-one or even large so that wave generation is strong and the physical implications are the most direct. Nonetheless, in interpreting these simulations, it is often possible to make contact with the theoretical developments valid for $\epsilon \ll 1$ summarised in the previous section.

If $\epsilon \gtrsim 1$ everywhere in a flow, the notion of balance is not well defined, and it is difficult to isolate a possible IGW component. But when $\epsilon \gtrsim 1$ only in localised regions, the separation between balanced motion and IGWs is possible away from these regions, which can then be regarded as localised sources of IGWs. In situations of this type, wave generation is not strictly speaking caused by balanced motion as discussed so far since the dynamics is not balanced at the generation site. A well studied example, which we now discuss, is the generation by fronts which arise near horizontal surfaces.

3.1 Surface-intensified flows

Balanced dynamics at horizontal surfaces — earth's surface and tropopause for the atmosphere, sea surface and floor for the ocean — is controlled by the advection of surface potential temperature (for the atmosphere) or buoyancy (for the ocean, see (5)). This dynamics is very different from the dynamics in the interior: whereas the cascade to small scales that characterises complex motion leads to Rossby numbers that are essentially scale independent in the interior,

the analogous cascade at the surface leads to Rossby numbers that increase as the flow scale decreases. This is demonstrated most clearly using the surface QG model (Blumen 1978, Juckes 1994, Held et al. 1995). The upshot is that large Rossby numbers $\epsilon \gtrsim 1$ appear locally in small-scale features such as fronts and filaments even though the flows are balanced at large scales; IGW generation is a consequence.

This type of generation has been studied in detail in PE versions of classical two-dimensional models of frontogenesis (Hoskins & Bretherton 1972, Williams 1967). In these models, the scale cascade is so rapid that the balanced evolution (described by the semi-geostrophic model) predicts a finite-time singularity. Two-dimensional numerical simulations show that the breakdown of balance associated with this singularity is accompanied by the generation of IGWs which are strong in the case of intense frontogenesis (Snyder et al. 1993, Griffiths & Reeder 1996, Reeder & Griffiths 1996, and references therein). A key observation is that the IGW amplitudes are the largest in frontogenesis scenarios that have the fastest evolution, consistent with the role played by the high-frequency content of the balanced signal discussed in section 2.4. Note that the finite-time singularity of the balanced motion invalidates the hypothesis of smooth time dependence that is critical for the exponential smallness of IGW generation. More recent work has examined fully three-dimensional flows: McWilliams et al. (2009) analyse the growth of three-dimensional unbalanced motion (not specifically IGWs) in a simple model of frontogenesis; Capet et al. (2008a,b) and Danioux et al. (2012) consider more complex, fully turbulent ocean flows, with the latter focusing on IGWs observed at depth.

3.2 Baroclinically unstable flows

Another mechanism of wave generation has been examined in simulations of baroclinically unstable flows by several authors, starting with O'Sullivan & Dunkerton (1995). This mechanism corresponds more closely to spontaneous generation by balanced flows as discussed in section 2.4. Unlike the frontal mechanism, it does not appear to rely on a cascade leading to $\epsilon \gtrsim 1$ at small scales and may be expected to persist for $\epsilon \ll 1$. In the simulations to date, however, ϵ is typically not small, so that the IGWs are readily detectable in fast fields such as horizontal divergence δ or vertical velocity.

The simulations of O'Sullivan & Dunkerton (1995) showed that near-balanced baroclinic life cycles are accompanied by the spontaneous generation of IGW wavepackets of relatively large horizontal scales (~ 500 km) and low frequencies ($\sim 1.5f$) near the exit of the jet that appears in the evolution. This result, initially controversial because of its apparent sensitivity to numerical resolution, has been confirmed and refined in several subsequent studies (Zhang 2004, Plougonven & Snyder 2005, Viùdez & Dritschel 2006, Plougonven & Snyder 2007). These demonstrate conclusively that IGWs are generated spontaneously in interior regions characterised by strong Lagrangian transience (IGWs are also generated near the surface, through the frontogenesis process described above). In most simulations, these regions neighbour the tropopause, so that the surface effects of section 3.1 may be relevant; however, IGWs also appear in the absence of a sharp tropopause (Viùdez & Dritschel 2006).

An example of IGW emission by baroclinic instability is shown in Fig. 6 (based on Plougonven & Snyder (2007)). The wave scales are relatively large, of the order of hundreds of kilometres, when the waves are generated, but they decrease

rapidly as the waves propagate through the sheared and strained background flow (Plougonven & Snyder 2005). This makes the simulation of wave propagation challenging, but the generation process is thought to be well resolved. The waves shown in the figure are generated near the tropopause; waves are also generated near the surface but appear later in the simulation. Numerical results such as these are in qualitative agreement with observations of IGWs thought to be generated spontaneously (e.g. Guest et al. 2000).

3.3 Dipolar flows

Particularly clear instances of spontaneous generation by balanced flows are provided by numerical studies of the propagation of vortex dipoles in the PEs. Two distinct but closely related examples demonstrate that IGWs accompany what is, in balanced approximations, the steady translation of a pair of counterrotating vortices. In the first example, the vortices correspond to PV anomalies in the fluid's interior (Viùdez 2006, 2007, 2008, Wang & Zhang 2009, Wang et al. 2009); in the second they correspond to anomalies of potential temperature at the lower boundary of a fluid with uniform PV (Snyder et al. 2007, 2009, Wang et al. 2009). In both cases, numerical simulations of the PEs are carried out with the steadily translating balanced solution as initial condition. This translating solution should persist in a slightly deformed way in higher-order balanced models. The PE simulations, however, reveal the presence of an IGW packet, translating at the speed of the dipole and located above (and below in the case of the interior dipole) it. Fig. 7, showing the vertical velocity for a surface-potential-temperature dipole, illustrates the spatial structure of the IGWs and suggests they have horizontal and vertical scales substantially shorter than those of the dipole. This can be un-

derstood from the dispersion relation (6): with U the dipole's translation speed, the frequency of the balanced motion, estimated as the Doppler-shift frequency $U\kappa$, matches the intrinsic frequency of IGWs provided that the horizontal and vertical wavenumbers satisfy $\kappa L = O(\epsilon^{-1})$ and $m/\kappa = O(1)$, where ϵ is based on the horizontal scale L of the dipole.[2]

Snyder et al. (2009) consider the mechanism of IGW generation in detail. They rule out the possibility of an unbalanced instability of the balanced flow (see section 4.2 below) and suggest instead that the IGWs are an integral part of the solution in the PEs. To provide support for this interpretation, they carry out a simulation of the linearised PEs forced by the difference between two time tendencies, that predicted by the QG approximation, and that predicted by the PEs. In essence, this amounts to considering the perturbative scheme (15)–(16) with $N = 1$ and with $\mathcal{L}$ the linearisation about the QG solution. The simulation produces an IGW pattern that is similar to that obtained in the PEs, although the amplitude is larger (see also Wang & Zhang 2009)). This suggests that the IGWs are a part of the full PE solution, which arises from the resonant projection of nonlinear balanced terms on the IGW operator as described in section 2.4. [3]

The IGWs radiated by vortex dipoles are analogous to IWGs generated by topography in small-Rossby-number flows. For these, the balanced response is also complemented by exponentially small IGWs (Muraki 2003). There are differences, however, including the fact that the flow over topography needs to be maintained against mountain drag, whereas in the dipole case the flow is un-

[2]As noted by Snyder et al. (2009), McIntyre (2009) and Vanneste (2008b), an analogous frequency matching is not possible for shallow-water flows because of the fixed vertical scale.

[3]For dipoles with piecewise-constant PV or potential temperature, the IGW amplitudes may not be exponentially small because the balanced fields are not completely smooth.

forced. Furthermore, although time-independent (in a translating frame of reference) in balanced approximations, the flow is presumably slowly evolving in the PEs because of IGW radiation. Another analogy can be drawn with solutions to a class of perturbed Koretweg–de Vries equations. These admit steadily translating solutions to all orders in the perturbation size, namely solitons with exponentially decaying tails, which turn out to be weakly radiating beyond all orders (Grimshaw & Joshi 1995, Pomeau et al. 1988). As in the dipole case, the complete solution (assuming no upstream influence) is not steadily translating because of the energy loss associated with radiation.

4 Analytical models

The numerical simulations discussed above provide ample evidence for spontaneous generation in near-balanced flows. However, the relation of these simulations with the theoretical arguments reviewed in sections 2.3–2.4 is not completely transparent, partly because the Rossby numbers used are at best moderately small. It is possible, however, to demonstrate the relevance of these arguments by considering much simpler flows which can be examined using the analytical methods of exponential asymptotics mentioned in section 2.4. These idealised flows reveal two types of mechanisms: the first is the generation of IGWs by transient balanced flow in a manner similar to the generation of fast oscillations in the LK model; the second is the instability of steady flows to unbalanced perturbations. In the first case, the amplitudes of the IGWs generated are exponentially small, whereas in the second case the growth rates of the instabilities are exponentially small.

4.1 Transient generation

The generation by transient balanced flows can be demonstrated by considering perturbations to simple background flows which depend linearly on space. Perturbations to such flows may be represented by plane waves with time-dependent wavevector and amplitudes (e.g. Craik & Allen 1992). The appeal of these solutions is broader since they also describe the dynamics of small-scale wavepackets propagating in more general background flows. This is the essence of the geometric-optics approach to fluid stability (e.g. Friedlander & Lipton-Lifschitz 2003) applied to rotating stratified flows in Guimbard & Leblanc (2006). It can be employed in the context of spontaneous generation to provide some intuition on the mechanism of transient generation (Aspden & Vanneste 2010).

The main idea is to add perturbations in the form of a small-scale wavepacket to a (possibly time-dependent) background solution of the PEs (2). Denoting the background velocity and buoyancy fields by $\boldsymbol{U}(\boldsymbol{x},t)$ and $B(\boldsymbol{x},t)$, we write

$$\boldsymbol{u} = \boldsymbol{U}(\boldsymbol{x},t) + \hat{\boldsymbol{u}}(\boldsymbol{x},t)\,\mathrm{e}^{\mathrm{i}\theta(\boldsymbol{x},t)/\delta} + \text{c.c.} \quad \text{and} \quad b = B(\boldsymbol{x},t) + \hat{b}(\boldsymbol{x},t)\mathrm{e}^{\mathrm{i}\theta(\boldsymbol{x},t)/\delta} + \text{c.c.},$$

where θ is a phase, $\delta \ll 1$ is a small parameter characterising the scale separation between large-scale background flow and small-scale perturbation, and c.c. denotes the complex conjugate of the previous term. The hatted variables are then expanded in powers of δ and the hypothesis of linearity $\hat{\boldsymbol{u}} \ll \boldsymbol{U}$ and $\hat{b} \ll B$ is made. Introducing into (2) gives $\partial_t\theta + \boldsymbol{U}\cdot\nabla\theta = 0$ to leading order. This implies that the phase θ can be obtained by solving the ODEs

$$\dot{\boldsymbol{x}} = \boldsymbol{U} \quad \text{and} \quad \dot{\boldsymbol{k}} = -\mathcal{D}^{\mathrm{T}}\boldsymbol{k}, \quad \text{where} \quad \mathcal{D} = \nabla\boldsymbol{U}, \tag{17}$$

for the position $\boldsymbol{x}(t)$ and wavevector $\boldsymbol{k}(t) = \nabla\theta(\boldsymbol{x}(t),t)$ of the wavepacket. Note that this corresponds to the standard WKB treatment of waves in slowly varying

media in the particular case of a dispersion relation $\omega = \boldsymbol{U}\cdot\boldsymbol{k}$ which neglects the intrinsic frequency compared to the Doppler-shift term.

The next order in δ yields equations for the hatted variables which are solved as ODEs along the characteristics in (17). Using incompressibility, $\boldsymbol{k}\cdot\hat{\boldsymbol{u}} = 0$, these can be reduced to equations for the PV

$$\hat{q} = (\boldsymbol{k}\times\hat{\boldsymbol{u}})\cdot(N^2\hat{\boldsymbol{z}} + \nabla B) + (\boldsymbol{\Omega}\cdot\boldsymbol{k})\hat{b},$$

with $\boldsymbol{\Omega} = f\hat{\boldsymbol{z}} + \nabla\times\boldsymbol{U}$ the absolute vorticity, vertical component of the vorticity, $\hat{\zeta} = \hat{\boldsymbol{z}}\cdot(\boldsymbol{k}\times\hat{\boldsymbol{u}})$, and vertical velocity $\hat{w} = \hat{\boldsymbol{z}}\cdot\hat{\boldsymbol{u}}$. These take the form

$$\begin{aligned}
\dot{\hat{q}} &= 0,\\
\dot{\hat{\zeta}} &= (\boldsymbol{\Omega}\cdot\boldsymbol{k})\hat{w} + \hat{\boldsymbol{z}}\cdot\mathcal{D}(\boldsymbol{k}\times\hat{\boldsymbol{u}}),\\
\dot{\hat{w}} &= -\frac{fm}{\kappa^2+m^2}\hat{\zeta} + \frac{\kappa^2}{\kappa^2+m^2}\hat{b} + \left(\frac{2m}{\kappa^2+m^2}\boldsymbol{k} - \hat{\boldsymbol{z}}\right)\cdot\mathcal{D}\hat{\boldsymbol{u}}.
\end{aligned}$$

This is a closed system of ODEs since $\hat{\boldsymbol{u}}$ and $\hat{b}$ can be expressed in terms of $\hat{q}$, $\hat{\zeta}$ and $\hat{w}$. It has been derived here in the geometric-optics context assuming $\delta \ll 1$ but can also be obtained without approximation if the background flow is exactly linear in $\boldsymbol{x}$, that is, if $\nabla\boldsymbol{U}$ and ∇B are constant. No assumption is made on ϵ, so the system may be used to examine spontaneous generation for $\epsilon = O(1)$. We focus here on the regime $\epsilon \ll 1$.

In this regime, $\boldsymbol{\Omega} = f\hat{\boldsymbol{z}} + O(\epsilon)$, $\nabla B = O(\epsilon)$ and the system reduces to

$$\dot{\hat{q}} = 0, \tag{18a}$$

$$\dot{\hat{\zeta}} - fm\hat{w} = O(\epsilon), \tag{18b}$$

$$\dot{\hat{w}} + \frac{1}{fm}\left(f^2 + \frac{(N^2-f^2)\kappa^2}{\kappa^2+m^2}\right)\hat{\zeta} = \frac{\kappa^2}{fm(\kappa^2+m^2)}\hat{q} + O(\epsilon). \tag{18c}$$

Its structure is closely analogous to that of the PEs (2) so that it provides a simple model to study spontaneous generation in a variety of flows. The bal-

anced flow is represented by both the background flow and the balanced part of the perturbation associated with $\hat{q}$. The effect of Lagrangian transience appear through the time dependence of the wavevector $\boldsymbol{k}$ governed by (17); introducing ϵ explicitly indicates that $\boldsymbol{k}$ evolves on the slow time scale: $\boldsymbol{k} = \boldsymbol{k}(\epsilon t)$.

The essence of the generation mechanisms described by (18) is best understood by writing a single ODE, say for the vertical vorticity $\hat{\zeta}$. With the substitution $t \mapsto \epsilon t$ corresponding to the use of the slow time scale, this takes the form

$$\epsilon^2 \left(\ddot{\hat{\zeta}} + b(t)\dot{\hat{\zeta}} \right) + \left(\omega^2(t) + O(\epsilon) \right) \hat{\zeta} = c(t)\hat{q}, \tag{19}$$

where the scaled IGW frequency $\omega(t) = \left(1 + (s^2 - 1)\kappa^2/(\kappa^2 + m^2)\right)^{1/2}$ and the functions $b(t)$ and $c(t)$ depend on time through $\kappa(t)$ and $m(t)$. The balanced part of $\hat{\zeta}$ corresponds to the inhomogeneous solution to this equation and is found as an asymptotic series, starting as

$$\hat{\zeta} = \frac{c(t)}{\omega^2(t)}\hat{q} + O(\epsilon), \tag{20}$$

in agreement with the general scheme described in section 2.4. The IGW part of the solution is obtained using a WKB approximation and reads

$$\hat{\zeta} = A_+ \mathrm{e}^{\mathrm{i}\int^t \omega(t')\,\mathrm{d}t'/\epsilon} \left(e_+(t) + O(\epsilon) \right) + A_- \mathrm{e}^{-\mathrm{i}\int^t \omega(t')\,\mathrm{d}t'/\epsilon} \left(e_-(t) + O(\epsilon) \right), \tag{21}$$

where $A_\pm$ are arbitrary constants and the functions $e_\pm(t)$ are expressed in terms of $\omega(t)$, $b(t)$ and the $O(\epsilon)$ terms in (19). Assuming smooth time dependence, the three components of the solution — balanced part and two IGWs with frequencies $\pm\omega$ — are uncoupled to all orders in ϵ. In particular, an initially balanced solution with $A_+ = A_- = 0$ remains balanced to all orders. Exponentially small effects lead to change in $A_\pm$. They give rise to two distinct mechanisms of spontaneous generation, both arising from the time dependence of the coefficients in (19).

The first is characterised by $\hat{q} \neq 0$ and $A_+ = A_- = 0$ at some initial time. The corresponding solution lies initially on an (optimally truncated) slow manifold but leaves it subsequently, when $A_\pm$ change to non-zero, exponentially small values. As explained in section 2.4, this occurs at times $t = \tau = \mathrm{Re}\, t_*$, where the singularities t_* and $\bar{t}_*$ of the balanced motion are seen from (20) to satisfy $\omega(t_*) = \omega(\bar{t}_*) = 0$. These times are (complex) turning points of (19) in whose neighbourhood both the expansion of the balanced solution and the WKB approximation break down. The physical interpretation is clear: when $\omega(t)$ is close to zero, there is no time scale separation, and balanced motion and IGW are fully coupled. An exponential-asymptotics analysis can be carried out to estimate the change in $A_\pm$ at $t = \tau$. The details are somewhat involved, but the dominant dependence of the change on ϵ is simple:

$$A_\pm \asymp \exp(-\alpha/\epsilon), \quad \text{where} \quad \alpha = |\mathrm{Im} \int_\tau^{t^*} \omega(t)\, \mathrm{d}t|, \tag{22}$$

where the symbol $\asymp$ indicates that a prefactor $\gamma\epsilon^\beta$ is ignored. This formula quantities the intuitive notion that spontaneous generation of IGWs is stronger in regions characterised by strong Lagrangian transience: indeed, the distance $|t_* - \tau|$ can be thought of as a Lagrangian time scale relevant to the generation process.

Vanneste & Yavneh (2004) applied matched asymptotics to obtain an estimate of the amplitude of the IGWs generated by this mechanism in the case of a horizontal, linear shear flow $\boldsymbol{U} = (\Lambda y, 0, 0)$ with $B = 0$ considered earlier by McWilliams & Yavneh (1998) (see also Ólafsdóttir et al. 2005). In this case, a plane wave with wavevector $\boldsymbol{k}(t) = (k, -\Lambda k t, m)$, where k and m are positive constants, is an exact nonlinear solution of the perturbation equations. The estimate shows that spontaneous generation is strongest for perturbations with

aspect ratio $m/k = N/f$, recognised as the natural aspect ratio of QG flows. The generation is also stronger in anticyclonic flows ($\Lambda > 0$) than in cyclonic flows ($\Lambda < 0$). Ólafsdóttir et al. (2008) took advantage of the fact that arbitrary linear perturbations can be represented as a superpositions of plane waves to obtain an approximation for the IGWs generated by a sheared ellipsoidal vortex.

More recently, Lott et al. (2010, 2012) used a different approach to estimate the amplitude of the IGWs associated with a PV perturbation in a vertical shear flow $\boldsymbol{U} = (\Lambda z, 0, 0)$. This approach emphasises the importance of inertial levels, that is, altitudes where the Doppler-shifted IGW frequency matches the Coriolis frequency f. A more complicated background flow, generated by a point-vortex dipole, is considered in Aspden & Vanneste (2010).

The second mechanism of IGW generation described by (19) is characterised by $\hat{q} = 0$ and $A_\pm \neq 0$ at $t = 0$ and represents a pure IGW pair. If (19) with $\hat{q} = 0$ predicts the growth of $A_\pm$, there is an instability which amplifies infinitesimal IGWs to finite amplitude. This can be interpreted as a mechanism of spontaneous generation even though it differs from the ones described so far by requiring a (small) initial imbalance. For flows with closed streamlines such that $\boldsymbol{k}(t)$ and hence the coefficients in (19) are periodic, some solutions are expected to grow through parametric instability (e.g. Bender & Orszag 1999) provided that some resonance conditions are satisfied. For $\epsilon \ll 1$, this translates into the existence of exponentially narrow regions in parameters space for which the solution grows at an exponentially small rate. This is demonstrated explicitly for the elliptical flow $\boldsymbol{U} = (ay, -bx, 0)$, where $a \neq b$ are constants, in Aspden & Vanneste (2009), following numerical work by Miyazaki (1993) and McWilliams & Yavneh (1998).

The growth of IGW amplitudes can also be expected in more complex flows,

e.g. leading to chaotic trajectories. Modelling the velocity gradient $\mathcal{D}$ along Lagrangian trajectories by stationary random processes, (19) is a linear ODE with multiplicative (stationary) noise; its solutions grow exponentially in the long-time limit at a rate given by the largest Lyapunov exponent. In the limit $\epsilon \to 0$, and under the assumption that the random processes are smooth, this Lyapunov exponent can be shown in simpler but similar ODEs to be exponentially small (A. M. Davie, personal communication). However, for non-smooth coefficients, with frequency spectra that decay algebraically rather than exponentially at high frequencies, a power-law dependence is obtained (see, e.g., Arnold et al. 1986). This raises the question of what the (Lagrangian, frequency) power spectrum of $\mathcal{D}$ is in realistic turbulent flows.

4.2 Unbalanced instabilities

Many papers assess how the stability properties of simple steady flows classically studied in balanced models change when considered in the context of the PEs. These papers identify unbalanced instabilities, that is, modes of instability that are filtered out in all balanced models and can therefore be interpreted as mechanisms of spontaneous generation. Eady's model of baroclinic instability, in particular, has been revisited by several authors using the PEs rather than QG equations, starting with Stone (1966, 1970) (see also Nakamura 1988). Similarly, Phillips's two-layer model has been reconsidered by Sakai (1989). These works uncover unbalanced instabilities associated with the coupling of either two unbalanced IGW or Kelvin modes, or of one unbalanced mode with a balanced mode. The focus was then on regimes with $\epsilon \gtrsim 1$, when these instabilities can dominate over balanced instabilities, but more recent work considers $\epsilon \ll 1$. In the case

of Eady's model, the growth rates of unbalanced instabilities can be shown to be exponentially small in ϵ (Molemaker et al. 2005), while for Phillips's model unbalanced instabilities only exist when ϵ sufficiently large (Gula et al. 2009). Exponential smallness is again a direct consequence of the fact that the instabilities are filtered out in optimally truncated balanced models. The difference between continuously stratified and layer models, on the other hand, is explained by frequency-matching arguments similar to those noted in section 3.3.

The existence of unbalanced instabilities is more striking in flows which, unlike those of the Eady and Phillips models, are stable in balanced approximations. Several examples of such flows have been studied using numerical or asymptotic methods. These include the rotating Taylor–Couette flow (Yavneh et al. 2001), uniform horizontal shear flow (Vanneste & Yavneh 2007), boundary current (McWilliams et al. 2004) and uniform vertical shear flow over a horizontal surface (Plougonven et al. 2005).

Among the mechanisms of unbalanced instability, several are examples of radiative instabilities, in which the unbalanced mode is an IGW radiating to infinity. Since most studies of radiative instabilities do not assume $\epsilon \ll 1$ (see Plougonven et al. 2005, however) but more often assume a small Froude number, we discuss them in the next section.

5 Small-Froude-number regime

As pointed out in section 2.1, strong stratification, corresponding to a Froude number $F \ll 1$ with $\epsilon = O(1)$, leads to a regime of time-scale separation between balanced motion and IGWs that differs from the small-Rossby-number regime $\epsilon \ll 1$ considered so far. To see this, rewrite the ratio (7) of (nonlinear) vortical

frequency to IGW frequency as

$$\frac{\omega_0}{\omega_\pm} \approx \frac{1}{(\epsilon^{-2} + F^{-2}(1-s^{-2})\gamma^2/(1+\gamma^2 H^2/L^2))^{1/2}},$$

where $\gamma = \kappa L/(mH)$ compares the aspect ratio κ/m of the IGWs with the aspect ratio H/L of the balanced flow. For $s = N/f > 1$, the condition $F \ll 1$ ensures that the time-scale separation holds, but only provided that $\gamma = O(1)$: IGWs with $\gamma = O(F)$ have a frequencies which match the balanced-mode frequency. As a result, the regime $F \ll 1$ differs drastically from the regime $\epsilon \ll 1$: not matter how small F is, IGWs with a small enough aspect ratio (usually because of large horizontal scales) can be excited resonantly by the balanced flow. Spontaneous generation of IGWs is therefore stronger for $F \ll 1$, with amplitudes or growth rates that scale like F^2.

In the shallow-water context, with $F = U/\sqrt{gH}$, where H is now the fluid depth, the frequency ratio takes the simpler form

$$\frac{\omega_0}{\omega_\pm} \approx \frac{1}{(\epsilon^{-2} + F^{-2}(\kappa L)^2)^{1/2}},$$

which makes clear that IGWs with long horizontal scales $\kappa^{-1} = O(F^{-1}L)$ can be excited resonantly. In the absence of rotation ($\epsilon = \infty$), the mechanism of this excitation is equivalent to the Lighthill radiation of sound by vortical motion in weakly compressible fluids (Lighthill 1952, Howe 2002), with F playing the role of the Mach number.[4] Ford (1994a) extended Lighthill's approach to the rotating case $\epsilon = O(1)$ and applied it to study IGWs emitted by a barotropically unstable flow. The approach, which relies on the spatial-scale separation between the localised balanced motion and the much longer IGWs, can be fully justified by

[4] A more distant analogue is the radiation of gravitational waves by Newtonian mass motion discovered by Einstein (see e.g Landau & Lifschitz 1975).

matched asymptotics (Crow 1970, Ford et al. 2000). It leads to a wave equation for the IGWs that is driven by nonlinear terms associated with the balanced motion; the quadrupolar nature of these terms ensures that, to leading order, the IGWs generated take the form of a quadrupole.

Several mechanisms of spontaneous generation relying on Lighthill radiation have been examined in shallow-water and continuously stratified fluids. These include emission by rotating elliptical vortices (Ford 1994c, Plougonven & Zeitlin 2002), by unstable jets (Ford 1994a, Sugimoto et al. 2008), and radiative instabilities of axisymmetric vortices (Ford 1994b, Schecter 2008, and references therein). Radiative instabilities can be interpreted as generalisations of the Broadbent & Moore (1979) instability of a Rankine vortex to include rotation and smooth PV profiles. At their heart is the coupling between a balanced mode (Rossby–Kelvin wave propagating along the vortex boundary for the Rankine vortex, a generalisation thereof for smooth vortices), and an IGW which radiates away from the vortex. Specifically, it is the backreaction of the radiation on the vortex that is responsible for the instability. The coupling occurs through a turning point, and the amplitude of the solution at this turning point controls the growth rate of the instability. Because the balanced mode decays algebraically away from the vortex when $F \ll 1$, the growth rate scales algebraically with F (like F^4 for the fastest growing mode). When $\epsilon \ll 1$, the spatial decay is exponential leading, to exponentially small growth rates (Dritschel & Vanneste 2006). Note that radiative instabilities have also been considered in the limit of large azimuthal or (for continuously stratified flows) axial wavenumber, where exponentially small growth rates are also obtained (e.g. Ford 1994c, Le Dizès & Billant 2009).

A comprehensive treatment of IGW generation for $F \ll 1$ is provided by the

shallow-water analysis of Ford et al. (2000). This paper carries out matched asymptotics to a large-enough power of F to capture not only the wave radiation mechanism in the manner of Lighthill (1952) and Crow (1970) but also its backreaction on the balanced motion. The result is a new model which, like balanced models, is governed entirely by the evolution of PV but which, unlike balanced models, accounts for the radiation of IGWs to infinity. A fundamental difference with balanced models is that the inversion relating the velocity field to the PV depends on the entire history of the PV (specifically the history of its quadrupole moment). This reflects the fact that the degrees of freedom associated with IGWs are not genuinely filtered out. Rather, they obey integrable equations (because their dynamics is linear) so that their form at any time t can be expressed explicitly in terms of the PV history $q(\boldsymbol{x}, t')$ for $0 \leq t' < t$.

Summary points

1. At large scales, the atmosphere and ocean can be regarded as two-time-scale systems, with the Rossby number $\epsilon \ll 1$ measuring the time-scale separation between slow balanced motion and fast IGWs.

2. Balance relationships, defining slow manifolds, and the corresponding IGW-filtering balanced models can in principle be constructed to accuracy $O(\epsilon^N)$ for N arbitrarily large. However, the perturbative schemes employed for this construction diverge as $N \to \infty$.

3. The divergence reflects the non-existence of exactly invariant slow manifolds and the inevitability of spontaneous IGW generation from near-balanced initial conditions. Because they can be eliminated to $O(\epsilon^N)$ for any N, the generation processes are exponentially weak in ϵ for balanced motion that

is smooth in time and space.

4. Numerical simulations have identified IGWs generated spontaneously in baroclinic life cycles, in dipolar flows, and in surface-intensified flows. These, together with analytical models, emphasise the importance of Lagrangian transience as an essential ingredient for spontaneous generation. The extreme sensitivity to ϵ typified by exponential smallness results in strongly non-uniform wave generation.

5. Spontaneous IGW generation can be captured analytically in simple models using the methods of exponential asymptotics; these highlight the key role played by complex-time singularities of the balanced motion.

6. There are crucial differences between the spontaneous generation in the PEs and in the shallow-water equations. These stem from the different dispersion relations, with the PEs allowing for IGWs with bounded frequencies as their horizontal scale decreases to 0.

7. A time-scale separation is possible for $\epsilon = O(1)$ if the Froude number is small, corresponding to strong stratification. The time-scale separation in this case is not complete: long IGWs are excited resonantly by slow balanced motion through the process of Lighthill radiation.

Future issues

1. Much of our theoretical understanding of spontaneous generation comes from ODE models, either heuristic low-order models or special solutions of the fluid equations. There is a need to extend the methods employed in order to treat PDE models and clarify how the full spatio-temporal structure

of the balanced motion controls IGW generation.

2. The problem of IGW generation by turbulent balanced flows is largely open. The power-law wavenumber and frequency spectra of turbulence suggest that the exponential dependence of IGW amplitudes on ϵ found in smooth flows is replaced by a power-law dependence.

3. Further numerical simulations are required to establish unambiguously the ϵ-dependence of IGW amplitudes in relatively simple flows such as baroclinic life cycles and in turbulent flows.

4. The contribution of spontaneous generation to processes driven by IGWs such as wave drag in the middle atmosphere and diapycnal mixing in the ocean remains to be estimated. Spontaneous generation may be expected to have a larger impact in the atmosphere than in the ocean where there are strong alternative sources of non-topographic IGWs (winds and tides).

5. Further connections need to be made between the mechanisms of spontaneous generation identified in theoretical studies, and observations of IGWs especially in the atmosphere.

6. IGWs are not the only form of unbalanced motion generated spontaneously: the order-one Rossby numbers that appear near horizontal boundaries, in particular, lead to strong unbalanced instabilities which may play an important role for the dissipation of the large-scale balanced circulation (Molemaker et al. 2010).

Acknowledgments. JV acknowledges the support of the UK Natural Environment Research Council, and thanks O. Bokhove, R. Plougonven and D. Wirosoetisno for comments on a draft version of this review. R. Plougonven

kindly provided figures 6 and 7.

References

Ablowitz MJ, Fokas AS. 1997. *Complex variables: introduction and applications.* Cambridge University Press

Allen JS. 1993. Iterated geostrophic intermediate models. *J. Phys. Oceanogr.* 23:2447–2461

Arnold L, Papanicolaou G, Wihstutz V. 1986. Asymptotic analysis of the Lyapunov exponent and rotation number of the random oscillator and applications. *SIAM J. Appl. Math.* 46:427–450

Aspden JM, Vanneste J. 2009. Elliptical instability of a rapidly rotating, strongly stratified fluid. *Phys. Fluids* 21:074104

Aspden JM, Vanneste J. 2010. Inertia-gravity-wave generation: a geometric-optic approach. In *IUTAM Symposium on Turbulence in the Atmosphere and Oceans*, ed. D Dritschel

Babin A, Mahalov A, Nicolaenko B. 2000. Fast singular oscillating limits and global regularity for the 3D primitive equations of geophysics. *M2AN* 34:201–222

Babin A, Mahalov A, Nicolaenko B. 2001. 3D Navier–Stokes and Euler equations with initial data characterized by uniformly large vorticity. *Indiana Univ. Math. J.* 50:1–35

Babin A, Mahalov A, Nicolaenko B. 2002. Fast singular oscillating limits of stably-stratified 3D Euler and Navier–Stokes equations, and ageostrophic wave fronts.

In *Large-scale atmosphere–ocean dynamics I: analytical methods and numerical models*, eds. J Norbury, I Roulstone. Cambridge University Press, 126–201

Balser W. 2000. *Formal power series and linear systems of meromorphic ordinary differential equations*. Universitext. New York: Springer-Verlag

Bender CM, Orszag SA. 1999. *Advanced mathematical methods for scientists and engineers*. Springer

Berry MV. 1989. Uniform asymptotic smoothing of Stokes's discontinuities. *Proc. R. Soc. Lond. A* 422:7–21

Berry MV. 2005. Universal oscillations of high derivatives. *Proc. R. Soc. Lond. A* 461:1735–1751

Blumen W. 1978. Uniform potential vorticity flow. Part I: theory of wave interactions and two-dimensional turbulence. *J. Atmos. Sci.* 35:421–432

Bokhove O, Shepherd TG. 1996. On Hamiltonian balanced dynamics and the slowest invariant manifold. *J. Atmos. Sci.* 53:276–297

Boyd JP. 1994. The slow manifold of a five-mode model. *J. Atmos. Sci.* 51:1057–1064

Broadbent EG, Moore DW. 1979. Acoustic destabilizatiopn of vortices. *Proc. R. Soc. Lond. A* 290:353–371

Camassa R. 1995. On the geometry of an atmospheric slow manifold. *Physica D* 84:357–397

Camassa R, Tin SK. 1996. The global geometry of the slow manifold in the Lorenz–Krishnamurthy model. *J. Atmos. Sci.* 53:3251–3264

Capet X, McWilliams JC, Molemaker MJ, Shchepetkin AF. 2008a. Mesoscale to

submesoscale transition in the california current system. Part I: Flow structure, eddy flux, and observational tests. *J. Phys. Oceanogr.* 38:29–43

Capet X, McWilliams JC, Molemaker MJ, Shchepetkin AF. 2008b. Mesoscale to submesoscale transition in the california current system. Part II: Frontal processes. *J. Phys. Oceanogr.* 38:44–64

Cotter CJ. 2004. *Model reduction for shallow water dynamics: balance, adiabatic invariance and subgrid modelling.* Ph.D. thesis, Imperial College London

Cotter CJ, Reich S. 2006. Semigeostrophic particle motion and exponentially accurate normal forms. *Multiscale Model. Sim.* 5:476–496

Craik ADD, Allen HR. 1992. The stability of three-dimensional time-periodic flows with spatially uniform strain rates. *J. Fluid Mech.* 234:613–627

Crow SC. 1970. Aerodynamic sound emission as a singular perturbation problem. *Stud. Appl. Math.* 49:21–44

Daley R. 1991. *Atmospheric data analysis.* Cambridge University Press

Danioux E, Vanneste J, Klein P, Sasaki H. 2012. Spontaneous inertia-gravity-wave generation by surface-intensified turbulence. *J. Fluid Mech.* In press

Dritschel DG, Vanneste J. 2006. Instability of a shallow-water potential-vorticity front. *J. Fluid Mech.* 561:237–254

Embid P, Majda AJ. 1996. Averaging over fast gravity waves for geophysical flows with arbitrary potential vorticity. *Comm. Partial Differential Equations* 21:619–658

Embid P, Majda AJ. 1998. Low Froude number limiting dynamics for stably stratified flow with small or finite Rossby number. *Geophys. Astrophys. Fluid Dyn.* 87:1–50

Ford R. 1994a. Gravity wave radiation from vortex trains in rotating shallow water. *J. Fluid Mech.* 281:81–118

Ford R. 1994b. The instability of an axisymmetric vortex with monotonic potential vorticity in rotating shallow water. *J. Fluid Mech.* 280:303–334

Ford R. 1994c. The response of a rotating ellipse of uniform potential vorticity to gravity wave radidation. *Phys. Fluids* 6:3694–3704

Ford R, McIntyre ME, Norton WA. 2000. Balance and the slow quasi-manifold: some explicit results. *J. Atmos. Sci.* 57:1236–1254

Fowler AC, Kember G. 1996. The Lorenz–Krishnamurthy slow manifold. *J. Atmos. Sci.* 53:1433–1437

Friedlander SJ, Lipton-Lifschitz A. 2003. Localized instabilities in fluids. In *Handbook of Mathematical Fluid Dynamics, vol. II*, eds. S Friedlander, D Serre. Elsevier Science, 289–353

Gelfreich V, Lerman L. 2002. Almost invariant elliptic manifold in a singularly perturbed Hamiltonian system. *Nonlinearity* 15:447–457

Gill AE. 1982. *Atmosphere-ocean dynamics*. Academic Press

Griffiths M, Reeder MJ. 1996. Stratospheric inertia-gravity waves generated in a numerical model of frontogenesis. I: Model solutions. *Q.J.R. Meteorol. Soc.* 122:1153–1174

Grimshaw R, Joshi N. 1995. Weakly nonlocal solitary waves in a singularly perturbed Korteweg-De Vries equation. *SIAM J. Appl. Maths.* 55:124–135

Guest F, Reeder M, Marks C, Karoly D. 2000. Inertia-gravity waves observed in the lower stratosphere over Macquarie Island. *J. Atmos. Sci.* 57:737–752

Guimbard D, Leblanc S. 2006. Local stability of the Abrashkin–Yakubovich family of vortices. *J. Fluid Mech.* 567:91–110

Gula J, Plougonven R, Zeitlin V. 2009. Ageostrophic instabilities of fronts in a channel in a stratified rotating fluid. *J. Fluid Mech.* 627:485–507

Hakim V. 1991. Computation of transcendental effects in growth problems: linear solvability conditions and nonlinear methods – the example of the geometric model. In *Asymptotics beyond all orders*, eds. H Segur, S Tanveer, H Levine. Plenum press, 15–28

Held I, Pierrehumbert R, Garner S, Swanson K. 1995. Surface quasi-geostrophic dynamics. *J. Fluid Mech.* 282:1–20

Hinch EJ. 1991. *Perturbation methods.* Cambridge University Press

Hoskins BJ, Bretherton FP. 1972. Atmospheric frontogenesis models: mathematical formulation and solutions. *J. Atmos. Sci.* 29:11–37

Howe MS. 2002. *Theory of vortex sound.* Cambridge University Press

Jacobs SJ. 1991. Existence of a slow manifold in a model system of equations. *J. Atmos. Sci.* 48:893–901

Juckes M. 1994. Quasi-geostrophic dynamics of the tropopause. *J. Atmos. Sci.* 51:2756–2768

Kreiss H. 1991. Problems with different time scales. *Acta Numerica* 1:101–139

Landau LD, Lifschitz EM. 1975. *The classical theory of fields.* Pergamon, 4th ed.

Le Dizès S, Billant P. 2009. Radiative instability in stratified vortices. *Phys. Fluids* 21:096602

Leith CE. 1980. Nonlinear normal mode initialization and quasi-geostrophic theory. *J. Atmos. Sci.* 37:958–968

Lighthill MJ. 1952. On sound generated aerodynamically, I. General theory. *Proc. R. Soc. Lond. A* 211:564–587

Lorenz EN. 1980. Attractor sets and quasi-geostrophic equilibrium. *J. Atmos. Sci.* 37:1685–1699

Lorenz EN. 1986. On the existence of a slow manifold. *J. Atmos. Sci.* 43:1547–1557

Lorenz EN. 1992. The slow manifold — What is it? *J. Atmos. Sci.* 49:2449–2451

Lorenz EN, Krishnamurthy V. 1987. On the nonexistence of a slow manifold. *J. Atmos. Sci.* 44:2940–2950

Lott F, Plougonven R, Vanneste J. 2010. Gravity waves generated by sheared potential-vorticity anomalies. *J. Atmos. Sci.* 67:157–170

Lott F, Plougonven R, Vanneste J. 2012. Gravity waves generated by sheared three-dimensional potential-vorticity anomalies. *J. Atmos. Sci.* Submitted

Lynch P. 2002. The swinging spring: a simple model for atmospheric balance. In *Large-scale atmosphere-ocean dynamics*, eds. I Roulstone, J Norbury, vol. II: Geometric methods and models. Cambridge University Press, 64–108

MacKay RS. 2004. Slow manifolds. In *Energy Localisation and Transfer*, eds. T Dauxois, A Litvak-Hinenzon, RS MacKay, A Spanoudaki. World Sci., 149–192

Majda AJ, Embid P. 1998. Averaging over fast gravity waves for geophysical flows with unbalanced initial data. *Theoret. Comput. Fluid. Dynamics* 11:155–169

Matthies K. 2001. Time-averaging under fast periodic forcing of parabolic partial differential equations: exponential estimates. *J. Diff. Eq.* 174:133–180

McIntyre ME. 2009. Spontaneous imbalance and hybrid vortex-gravity structures. *J. Atmos. Sci.* 66:1315–1326

McIntyre ME, Norton WA. 2000. Potential vorticity inversion on a hemisphere. *J. Atmos. Sci.* 57:1214–1235. Corrigendum: **58**, 949-949 (2001)

McIntyre ME, Roulstone I. 2002. Are there higher-accuracy analogues of semi-geostrophic theory? In *Large-scale atmosphere-ocean dynamics*, eds. I Roulstone, J Norbury, vol. II: Geometric methods and models. Cambridge University Press, 301–364

McWilliams JC, Molemaker MJ, Olafsdottir EI. 2009. Linear fluctuation growth during frontogenesis. *J. Phys. Oceanogr.* 39:3111–3129

McWilliams JC, Molemaker MJ, Yavneh I. 2004. Ageostrophic, anticyclonic instability of a geostrophic, barotropic boundary current. *Phys. Fluids* 16:3720–3725

McWilliams JC, Yavneh I. 1998. Fluctuation growth and instability associated with a singularity of the balance equations. *Phys. Fluids* 10:2592–2596

Miyazaki T. 1993. Elliptical instability in a stably stratified rotating fluid. *Phys. Fluids* A 5:2702–2709

Mohebalhojeh AR, Dritschel DG. 2001. Hierarchies of balance conditions for the f-plane shallow water equations. *J. Atmos. Sci.* 58:2411–2426

Mohebalhojeh AR, McIntyre ME. 2007. Local mass conservation and velocity splitting in pv-based balanced models. Part i: the hyperbalance equations. *J. Atmos. Sci.* 64:1782–1793

Molemaker MJ, McWilliams JC, Capet X. 2010. Balanced and unbalanced routes to dissipation in an equilibrated Eady flow. *J. Fluid Mech.* 654:35–63

Molemaker MJ, McWilliams JC, Yavneh I. 2005. Baroclinic instability and loss of balance. *J. Phys. Oceanogr.* 35:1505–1517

Muraki DJ. 2003. Revisiting Queney's flow over a mesoscale ridge Unpublished

Muraki DJ, Snyder C, Rotunno R. 1999. The next-order correction to quasi-geostrophic theory. *J. Atmos. Sci.* 56:1547–1559

Nakamura N. 1988. Scale selection of baroclinic instability—Effects of stratification and nongeostrophy. *J. Atmos. Sci.* 45:3253–3267

Neishtadt AI. 1982. On the accuracy of conservation of adiabatic invariants. *J. Appl. Math. Mech.* 45:58–63

Nekhoroshev NN. 1977. An exponential estimate for the stability time of hamiltonian systems close to integrable ones. *Russ. Math. Surveys* 32:1–65

Ólafsdóttir EI, Olde Daalhuis AB, Vanneste J. 2005. Stokes-multiplier expansion in an inhomogeneous differential equation with a small parameter. *Proc. R. Soc. Lond. A* 461:2243–2256

Ólafsdóttir EI, Olde Daalhuis AB, Vanneste J. 2008. Inertia-gravity-wave radiation by a sheared vortex. *J. Fluid Mech.* 596:169–189

Olde Daalhuis AB, Chapman SJ, King JR, Ockendon JR, Tew RH. 1995. Stokes phenomenon and matched asymptotic expansions. *SIAM J. Appl. Math.* 55:1469–1483

O'Sullivan D, Dunkerton TJ. 1995. Generation of inertia-gravity waves in a simulated life-cycle of baroclinic instability. *J. Atmos. Sci.* 52:3695–3716

Plougonven R, Muraki DJ, Snyder C. 2005. A baroclinic instability that couples balanced motions and gravity waves. *J. Atmos. Sci.* 62:1545–1559

Plougonven R, Snyder C. 2005. Gravity waves excited by jets: propagation versus generation. *Geophys. Res. Lett.* 32:L18802

Plougonven R, Snyder C. 2007. Inertia-gravity waves spontaneously excited by jets and fronts. Part I: different baroclinic life cycles. *J. Atmos. Sci.* 64:2502–2520

Plougonven R, Zeitlin V. 2002. Internal gravity wave emission from a pancake vortex: an example of wave-vortex interaction in strongly stratified flows. *Phys. Fluids.* 14:1259–1268

Pomeau Y, Ramani A, Grammaticos B. 1988. Structural stability of the Korteweg-De Vries solitions under a singular perturbation. *Physica D* 31:227–134

Reeder MJ, Griffiths M. 1996. Stratospheric inertia-gravity waves generated in a numerical model of frontogenesis. Part II: Wave sources, generation mechanisms and momentum fluxes. *Q.J.R. Meteorol. Soc.* 122:1175–1195

Reznik GM, Zeitlin V, Ben Jelloul M. 2001. Nonlinear theory of geostrophic adjustment. Part 1. Rotating shallow-water model. *J. Fluid Mech.* 445:93–120

Rossby CG. 1937. On the mutual adjustment of pressure and velocity distributions in certain simple current systems,. *J. Mar. Res.* 21:15–28

Sakai S. 1989. Rossby-Kelvin instability: a new type of ageostrophic instability caused by a resonance between Rossby waves and gravity waves. *J. Fluid Mech.* 202:149–176

Salmon R. 1983. Practical use of Hamilton's principle. *J. Fluid Mech.* 132:431–444

Salmon R. 1985. New equations for nearly geostrophic flow. *J. Fluid Mech.* 153:461–477

Saujani S, Shepherd TG. 2006. A unified theory of balance in the extratropics. *J. Fluid Mech.* 569:447–464

Schecter DA. 2008. The spontaneous imbalance of an atmospheric vortex at high rossby number. *J. Atmos. Sci.* 65:2498–2521

Segur H, Tanveer S, Levine H, eds. 1991. *Asymptotics beyond all orders*, New York. Plenum Press

Snyder C, Muraki D, Plougonven R, Zhang F. 2007. Inertia-gravity waves generated within a dipole vortex. *J. Atmos. Sci.* 64:4417–4431

Snyder C, Plougonven R, Muraki D. 2009. Mechanism for spontaneous gravity wave generation within a dipole vortex. *J. Atmos. Sci.* 66:3464–3478

Snyder C, Skamarock W, Rotunno R. 1993. Frontal dynamics near and following frontal collapse. *J. Atmos. Sci.* 50:3194–3211

Stone PH. 1966. On non-geostrophic baroclinic instability. *J. Atmos. Sci.* 23:390–400

Stone PH. 1970. On non-geostrophic baroclinic instability. Part II. *J. Atmos. Sci.* 27:721–726

Sugimoto N, Ishioka K, Ishii K. 2008. Parameter sweep experiments on spontaneous gravity wave radiation from unsteady rotational flow in an f-plane shallow water system. *J. Atmos. Sci.* 65:235–249

Temam R, Wirosoetisno D. 2007. Exponentially accurate approximations for the primitive equations of the ocean. *Discr. Cont. Dyn. Sys. B* 7:425–440

Temam R, Wirosoetisno D. 2010. Stability of the slow manifold in the primitive equations. *SIAM J. Math. Anal.* 42:427–458

Temam R, Wirosoetisno D. 2011. Slow manifolds and invariant sets of the primitive equations. *J. Atmos. Sci.* 68:675–682

Vallis GK. 2006. *Atmospheric and oceanic fluid dynamics: fundamentals and large-scale circulation.* Cambridge University Press

Vanneste J. 2004. Inertia-gravity-wave generation by balanced motion: revisiting the Lorenz-Krishnamurthy model. *J. Atmos. Sci.* 61:224–234

Vanneste J. 2008a. Asymptotics of a slow manifold. *SIAM J. Appl. Dynam. Syst.* 7:1163–1190

Vanneste J. 2008b. Exponential smallness of inertia-gravity-wave generation at small Rossby number. *J. Atmos. Sci.* 65:1622–1637

Vanneste J, Yavneh I. 2004. Exponentially small inertia-gravity waves and the breakdown of quasi-geostrophic balance. *J. Atmos. Sci.* 61:211–223

Vanneste J, Yavneh I. 2007. Unbalanced instabilities of rapidly rotating stratified shear flows. *J. Fluid Mech.* 584:373–396

Vautard R, Legras B. 1986. Invariant manifolds, quasi-geostrophy and initialization. *J. Atmos. Sci.* 43:565–584

Viùdez A. 2006. Spiral patterns of inertia-gravity waves in geophysical flows. *J. Fluid Mech.* 562:73–82

Viùdez A. 2007. The origin of the stationary frontal wave packet spontaneously generated in rotating stratified vortex dipole. *J. Fluid Mech.* 593:359–383

Viùdez A. 2008. The stationary frontal wave packet spontaneously generated in mesoscale dipoles. *J. Phys. Oceanogr.* 38:243256

Viùdez A, Dritschel DG. 2006. Spontaneous emission of inertia-gravity wave packets by balanced geophysical flows. *J. Fluid Mech.* 553:107–117

Wang S, Zhang F. 2009. Source of gravity waves within a vortex-dipole jet revealed by a linear model. *J. Atmos. Sci.* 67:1438–1455

Wang S, Zhang F, Snyder C. 2009. Generation and propagation of inertia-gravity waves from vortex dipoles and jets. *J. Atmos. Sci.* 66:1294–1314

Warn T. 1986. Statistical mechanical equilibria of the shallow water equations. *Tellus* 38A:1–11

Warn T. 1997. Nonlinear balance and quasi-geostrophic sets. *Atmos.-Ocean* 35:135–145

Warn T, Bokhove O, Shepherd TG, Vallis GK. 1995. Rossby number expansions, slaving principles, and balance dynamics. *Quart. J. R. Met. Soc.* 121:723–739

Warn T, Menard R. 1986. Nonlinear balance and gravity-inertial wave saturation in a simple atmospheric model. *Tellus* 38A:285–294

Williams RT. 1967. Atmospheric frontogenesis: a numerical experiment. *J. Atmos. Sci.* 24:627–641

Wirosoetisno D. 2004. Exponentially accurate balance dynamics. *Adv. Diff. Eq.* 9:177–196

Wirosoetisno D, Shepherd TG, Temam RM. 2002. Free gravity waves and balanced dynamics. *J. Atmos. Sci.* 59:3382–3398

Yavneh I, McWilliams JC, Molemaker MJ. 2001. Non-axisymmetric instability of centrifugally-stable stratified Taylor-Couette flow. *J. Fluid Mech.* 448:1–21

Yuan L, Hamilton K. 1994. Equilibrium dynamics of forced-dissipative f-plane shallow-water system. *J. Fluid Mech.* 280:369–394

Zeitlin V, Reznik GM, Ben Jelloul M. 2003. Nonlinear theory of geostrophic adjustment. Part 2. Two-layer and continuously stratified primitive equations. *J. Fluid Mech.* 491:207–228

Zhang F. 2004. Generation of mesoscale gravity waves in upper-tropospheric jet-front systems. *J. Atmos. Sci.* 61:440–457

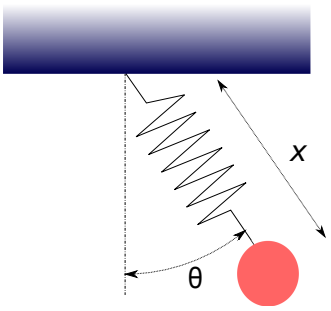

Figure 1: The elastic pendulum is a simple example of a two-time-scale dynamical system when the natural frequency of vibration of the spring is large compared to the natural frequency of oscillation of the pendulum. The slow angle θ and the fast extension x serve as analogues for the slow balanced motion and fast IGWs of geophysical fluids.

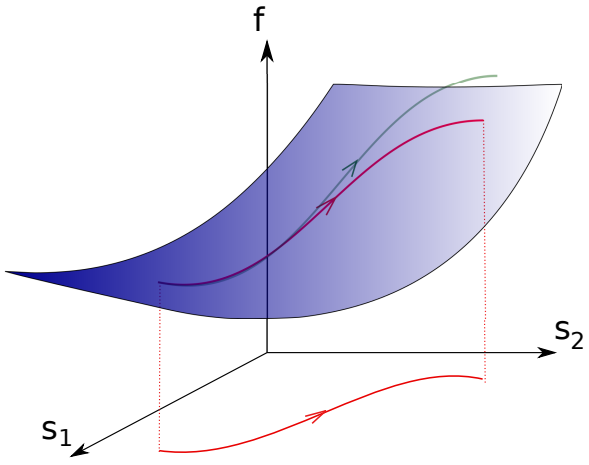

Figure 2: Schematic of a slow manifold: the (blue) manifold is approximately invariant so that an exact trajectory (green curve) starting on it remains close for some time. Also shown is a solution of the corresponding balanced model (14) in the slow $(\mathsf{s}_1, \mathsf{s}_2)$-plane and its lift to the slow manifold (red curves).

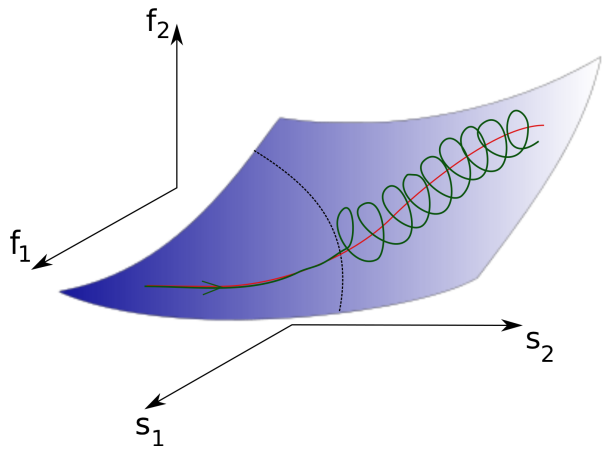

Figure 3: Spontaneous generation of fast oscillations: an exact trajectory (green curve) starting on a highly accurate slow manifold develops fast oscillations and hence deviates from the balanced approximation (red curve). The oscillations are exponentially small and appear suddenly when the trajectory crosses a Stokes line (dotted curve).

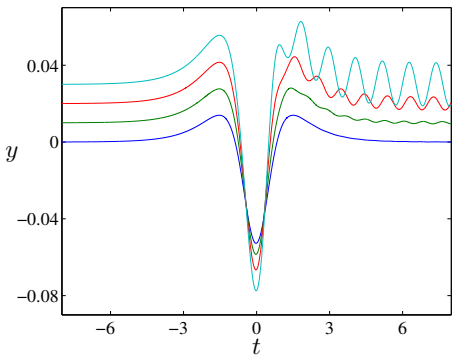

Figure 4: Evolution of the fast variable y in the LK model for $\epsilon = 0.1$ (blue), 0.125 (green), 0.15 (red) and 0.175 (cyan) and $b = 0.5$ (the four curves are offset in the vertical for clarity). The balanced trajectory is asymptotic to an unstable equilibrium (analogous to the upright position of a pendulum) for $t \to \pm\infty$.

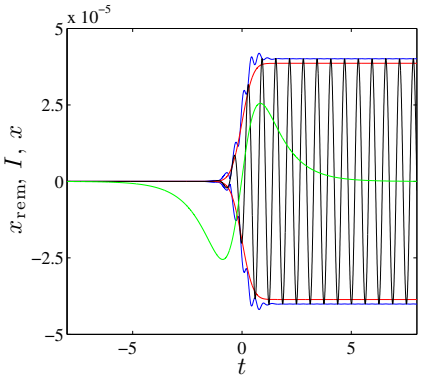

Figure 5: Unbalanced contribution to the fast variable x for the LK model with $\epsilon = 0.1$ and $b = 0.5$ (black curve). This is estimated numerically as $x_{\mathrm{rem}} = x - x_{\mathrm{bal}}$ using a balance relation for x_{bal} accurate to $O(\epsilon^{12})$. The amplitude of fast oscillations $I = \sqrt{x_{\mathrm{rem}}^2 + y_{\mathrm{rem}}^2}$ (blue curve) is compared with its asymptotic approximation in terms of an error function (red curve). The evolution of the total x, scaled by 10^4, is also indicated (green curve).

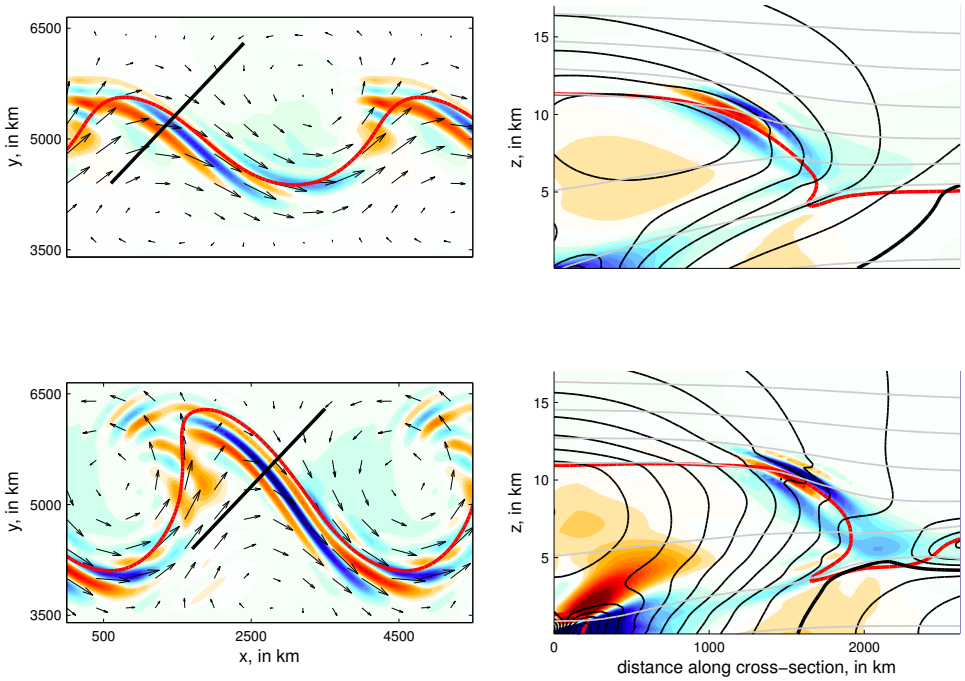

Figure 6: IGW generation by baroclinic instability. Cross sections of the divergence δ (colours, in the range $\pm 0.4f$) at days 7 (upper) and 8 (lower) of an idealised baroclinic life cycle simulation, as in Plougonven & Snyder (2005, 2007). Left: horizontal cross-sections of δ at $z = 11$ km, also showing horizontal velocity (arrows) and the tropopause (thick red line corresponding to 2PV units at $z = 9$ km). Right: vertical cross-sections of δ, taken along the black segment in the left panel, showing also the horizontal velocity (black contours) and the potential temperature (grey contours). (Figure courtesy of R. Plougonven.)

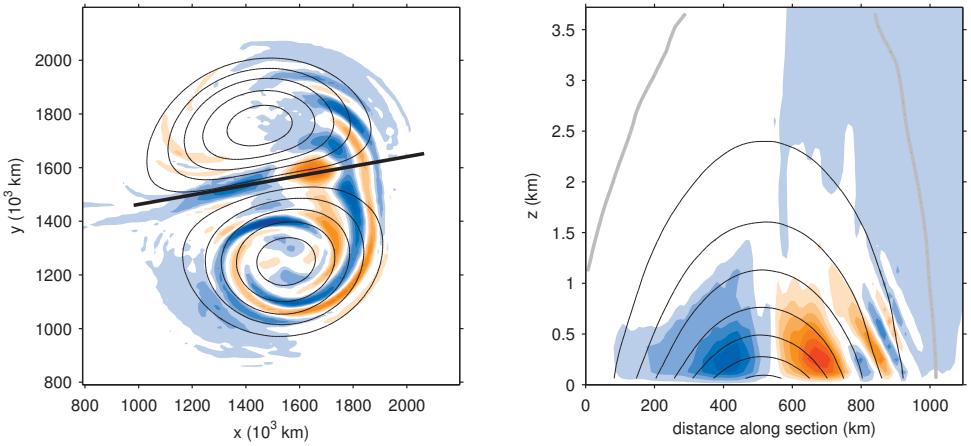

Figure 7: IGW generation by a surface-potential-temperature dipole. Cross sections of the vertical velocity w (coloured). Left: horizontal cross-section of w at $z = 125$ m, also showing the potential temperature (at $z = 62.5$ m, black contours). Right: vertical section of w taken along the black segment on the left panel, also showing the horizontal velocity parallel to the section (black contours). (From Snyder et al. (2007). ©American Meteorological Society. Reprinted with permission.)